

# Large Deviation implies First and Second Laws of Thermodynamics

Hiroyasu Tajima,<sup>1</sup> Eyuri Wakakuwa,<sup>2</sup> and Tomohiro Ogawa<sup>2</sup>

<sup>1</sup> Center for Emergent Matter Science (CEMS), RIKEN, Wako, Saitama 351-0198 Japan

<sup>2</sup> Graduate School of Information Systems, The University of Electro-Communications, Chofu, Japan

To reconstruct thermodynamics based on the microscopic laws is one of the most important unfulfilled goals of statistical physics. Here, we show that the first law and the second law for adiabatic processes are derived from an assumption that “probability distributions of energy in Gibbs states satisfy large deviation”, which is widely accepted as a property of thermodynamic equilibrium states. We define an adiabatic transformation as a probabilistic mixture of the energy-preserving unitary transformations on the many-body systems and the work storage. As the second law, we show that an adiabatic transformation from a set of Gibbs states to another set of Gibbs states is possible if and only if the regularized von Neumann entropy becomes large. As the first law, we show that the energy loss of the thermodynamic systems during the adiabatic transformation is stored in the work storage as “work,” in the following meaning; (i) the energy of the work storage takes certain values macroscopically, in the initial state and the final state. (ii) the entropy of the work storage in the final state is macroscopically equal to the entropy of the initial state. As corollaries, our results give the principle of maximum work and the first law for the isothermal processes.

## I. INTRODUCTION

Statistical mechanics was born to reconstruct thermodynamics from the microscopic mechanical laws [1]. It has achieved great results and succeeded in elucidating various nonequilibrium phenomena beyond thermodynamic region [2]. However, the original goal of statistical mechanics, “Reconstructing thermodynamics from microscopic mechanical laws” has not been fully accomplished [3, 4]. In thermodynamics, whether macroscopic state transformation is possible or not is completely determined by the values of the thermodynamic functions of the initial and final states of the transformation [3, 4]. In the case of adiabatic processes, a combination of the equilibrium states  $(T_1, V_1, N_1) \times \dots \times (T_m, V_m, N_m)$  can be adiabatically transformed to another combination  $(T'_1, V'_1, N_1) \times \dots \times (T'_m, V'_m, N_m)$  if and only if the following inequality holds:

$$\Delta S_S \geq 0, \quad (1)$$

where  $\Delta S_S$  is the difference of the thermodynamic entropy of the combinations  $(T_1, V_1, N_1) \times \dots \times (T_m, V_m, N_m)$  and  $(T'_1, V'_1, N_1) \times \dots \times (T'_m, V'_m, N_m)$ , i.e.,  $\Delta S_S := \sum_k S(T'_k, V'_k, N_k) - S(T_k, V_k, N_k)$ . The energy lost from the thermodynamic systems during the adiabatic transformation is stored in the external work storage in a special form called “work:”

$$W = -\Delta U_S \quad (2)$$

These are the first law and the second law of thermodynamics. They are the two principles which form the basis of thermodynamics.

In order to derive these two laws from microscopic mechanical laws, much effort has been done in the field of statistical mechanics [2, 5–12]. However, the reconstruction of these laws is still not fully achieved. Regarding

the second law, although these results have been successfully demonstrated to show the direct part “the inequality holds if the conversion is possible,” the converse part “the conversion is possible if the inequality holds” has not been shown yet. Regarding the first law, although numerous studies [13, 14] try to distinguish a special energy form “work” from another form “heat” statistically mechanically, a complete conclusion has not been obtained yet.

Recently, the above problems in reconstruction of the two laws are rapidly being solved by the approach from quantum information theory [15–30]. These results give many necessary and sufficient conditions for the possibility of the transformation caused by the interaction with the heat baths. The conditions can be interpreted as extensions of the second law into nanoscale, and are identical with the second law of thermodynamics when the size of the system is infinitely large. In spite of the splendid success, however, the following open problems remain:

1. The results are based on strong assumptions about thermodynamic systems and heat baths, e.g., the i.i.d. feature and/or the number of degeneracy. The assumptions are not necessarily satisfied by actual thermodynamic systems. Ref. [30] shows that the second law in  $(U, V, N)$  expression is derived from a weak assumption that the regularized Boltzmann entropy exists, but the second law in  $(T, V, N)$  expression is still remain.
2. The analysis is mainly limited to cases where the baths are sufficiently large compared to the system, so that the temperatures of baths are almost unchanged during the thermal processes. In the adiabatic transformation without such large heat baths, the temperatures of all systems greatly change. Understanding how such temperature changes occur is still limited.

3. There is no consensus about the statistical mechanical definitions of “work” and “heat.” Some of the above studies [15–20, 23, 24] also approach to this problem, by employing the notion of the single-shot work extraction, in which the work extraction is defined as a deterministic translation from the ground pure state to the excited pure state in a two-level system. It defines a work-like energy transfer in the quantum scale well. However, it is not clear whether these formulations completely capture the notion of “work” in thermodynamics.

In this paper, we tackle these problems. We show that the first law and the second law are derived from an assumption that “probability distributions of energy in Gibbs states satisfy large deviation”, which is widely accepted as a property of thermodynamic equilibrium states. We treat thermodynamic systems and an external work storage as quantum systems, and define an adiabatic transformation as a probabilistic mixture of the energy-preserving unitary transformations on the thermodynamic systems and the work storage. As the second law, we show that an adiabatic transformation from a combination of Gibbs states corresponding to  $(T_1, V_1, N_1) \times \dots \times (T_m, V_m, N_m)$  to another combination of Gibbs states corresponding to  $(T'_1, V'_1, N'_1) \times \dots \times (T'_m, V'_m, N'_m)$  is possible if and only if the regularized von Neumann entropy becomes large:

$$\Delta \tilde{S}_S \geq 0, \quad (3)$$

where  $\Delta \tilde{S}_S$  is the difference of the regularized von Neumann entropy defined for the macroscopic parameters  $(T_1, V_1, N_1) \times \dots \times (T_m, V_m, N_m)$  and  $(T'_1, V'_1, N'_1) \times \dots \times (T'_m, V'_m, N'_m)$ ;  $\Delta \tilde{S}_S := \sum_k \tilde{S}(T'_k, V'_k, N'_k) - S(T_k, V_k, N_k)$ . As the first law, we show that the energy lost from the thermodynamic systems during the adiabatic transformation is stored in the work storage as “work,” in the following meaning; (i) the energy of the work storage takes certain values macroscopically, in the initial state and the final state. (ii) the entropy of the work storage in the final state is macroscopically equal to the entropy of the initial state.

As corollaries, our results give other forms of the first law and the second law. As an example, we give the maximum work principle  $\tilde{W} \leq -\Delta \tilde{F}_S$  and the first law  $\tilde{W} = -\Delta \tilde{U}_S + \tilde{Q}$  in the isothermal processes.

## II. FORMULATION

In this section, we introduce the formulation treated in the present paper. The system we deal with is divided into two parts, i.e., the internal system  $I$  and the external work storage  $E_W$ . The internal system is the composite system of the thermodynamic systems and the control system. We firstly explain the internal system, secondly explain the external work storage, and finally explain the dynamics on  $IE_W$ .

*Internal system;* We consider the internal system  $I$  as the composite system of the thermodynamic systems  $S_1, \dots, S_m$  and the control system  $C$ . We refer to the Hilbert spaces of  $S_1, \dots, S_m$  and  $C$  as  $\mathcal{H}_{S_1}^{(n)}, \dots, \mathcal{H}_{S_m}^{(n)}$  and  $\mathcal{H}_C$ , respectively. Here,  $n$  is the scaling parameter, and the macroscopic limit is defined as the limit of  $n \rightarrow \infty$ . We consider each  $S_k$  as a composite system of  $n$  subsystems, whose dimension is  $d_k$ . For example, when  $S_k$  is a composite system of  $n$   $d_k$ -level spins,  $d_k$  is finite, and the dimension of  $\mathcal{H}_{S_k}^{(n)}$  is  $d_k^n$ . When  $S_k$  is a composite system of  $n$  harmonic oscillators,  $d_k$  is infinite, and the dimension of  $\mathcal{H}_{S_k}^{(n)}$  is also infinite. For the simplicity, we refer to the composite system of  $S_1 \dots S_m$  as  $S$ .

The Hamiltonian of the internal system  $I$  is described as follows:

$$H_I^{(n)} := \sum_{\vec{\lambda}=(\lambda_1, \dots, \lambda_m)} (H_{1|\lambda_1}^{(n)} + \dots + H_{m|\lambda_m}^{(n)}) \otimes |\vec{\lambda}\rangle \langle \vec{\lambda}|_C, \quad (4)$$

where  $\vec{\lambda} := (\lambda_1, \dots, \lambda_m)$  represents the set of the control parameters, e.g., the strength of the magnetic field or the position of the piston. These parameters  $\vec{\lambda}$  are registered in a finite-dimensional system  $C$  as the state  $|\vec{\lambda}\rangle$  on  $\mathcal{H}_C$ . The operator  $H_{l|\lambda_l}^{(n)}$  is the Hamiltonian on  $\mathcal{H}_{S_l}^{(n)}$ . It is given as a function of the control parameter  $\lambda_l$ . By using Hamiltonian  $H_I^{(n)}$ , we can effectively change the Hamiltonians of  $S_1, \dots, S_m$  by changing the state of  $C$  [26]. For the simplicity, we use the abbreviation  $H_{S|\vec{\lambda}}^{(n)} := H_{1|\lambda_1}^{(n)} + \dots + H_{m|\lambda_m}^{(n)}$ .

As the equilibrium states of  $S_1, \dots, S_m$ , we employ Gibbs states. We describe the Gibbs state of  $S_l$  with the control parameter  $\lambda_l$  and the inverse temperature  $\beta_l$  as follows:

$$\rho_{\beta_l|\lambda_l}^{(n)} := \frac{\exp[-\beta_l H_{l|\lambda_l}^{(n)}]}{\text{Tr}[\exp[-\beta_l H_{l|\lambda_l}^{(n)}]]} \quad (5)$$

We assume that the initial state  $\rho_I^{(n)}$  of  $I$  is as follows:

$$\rho_I^{(n)} = \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle \langle \vec{\lambda}|_C, \quad (6)$$

where we use the abbreviation  $\rho_{\vec{\beta}|\vec{\lambda}}^{(n)} := \rho_{\beta_1|\lambda_1}^{(n)} \otimes \dots \otimes \rho_{\beta_m|\lambda_m}^{(n)}$ .

Finally, we require that when the thermodynamic systems  $S_1, \dots, S_m$  are in Gibbs states, they satisfy the following large deviation principle of energy. That is, for an arbitrary positive number  $x$ , the following approximations hold:

$$\text{Tr}[\rho_{\beta_l|\lambda_l}^{(n)} \Pi_{l, \geq n(\tilde{U}_l(\beta_l, \lambda_l) + x)}^{(n)}] \approx \exp[-n I_l(x|\beta_l, \lambda_l)]. \quad (7)$$

$$\text{Tr}[\rho_{\beta_l|\lambda_l}^{(n)} \Pi_{l, \leq n(\tilde{U}_l(\beta_l, \lambda_l) - x)}^{(n)}] \approx \exp[-n I_l(-x|\beta_l, \lambda_l)]. \quad (8)$$

where  $\tilde{U}_l(\beta_l, \lambda_l)$  is the regularized internal energy

$$\tilde{U}_l(\beta_l, \lambda_l) := \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr}[\rho_{\beta_l|\lambda_l}^{(n)} H_{l|\lambda_l}^{(n)}], \quad (9)$$

and where  $\Pi_{l, \geq na}^{(n)}$  and  $\Pi_{l, \leq na}^{(n)}$  are the projection to the subspace of  $\mathcal{H}_{S_l}^{(n)}$  which is spanned by the energy eigenvectors of  $H_{l|\lambda_l}^{(n)}$  whose eigenvalues are larger than or equal to  $na$ , and are lower than or equal to  $na$ , respectively. Moreover,  $I_l(x|\beta_l, \lambda_l)$  is a convex function which becomes 0 only when  $x = 0$ .

For Gibbs state satisfying the above assumption, the regularized Helmholtz free energy and the regularized entropy are also defined as follows:

$$\tilde{F}_l(\beta_l, \lambda_l) := \lim_{n \rightarrow \infty} \frac{-\log \text{Tr}[\exp[-\beta_l H_{l|\lambda_l}^{(n)}]]}{n\beta_l} \quad (10)$$

$$\tilde{S}_l(\beta_l, \lambda_l) := \lim_{n \rightarrow \infty} \frac{-\text{Tr}[\rho_{\beta_l|\lambda_l}^{(n)} \log \rho_{\beta_l|\lambda_l}^{(n)}]}{n} \quad (11)$$

These two quantities satisfy the following relation:

$$\tilde{S}_l(\beta_l, \lambda_l) = \beta_l(\tilde{U}_l(\beta_l, \lambda_l) - \tilde{F}_l(\beta_l, \lambda_l)). \quad (12)$$

We assume that  $\tilde{U}$  and  $\tilde{S}$  are continuous and decreasing functions of  $\beta$ . This assumption implies that the phase transition does not occur. Even when dealing with a system with a phase transition, if we limit the temperature range to be handled to a region without phase transition, the results of this paper is perfectly applicable.

*Work storage:* Next, we formulate the work storage  $E_W$  which stores the energy extracted from  $I$ . As the preparation, we clarify the concept of “work” treated in this paper. Intuitively, it can be said that the work extraction satisfies at least the following two conditions:

1. Before and after the work extraction, the energy of the work storage must have definite values macroscopically.
2. The entropy of the work storage must not change macroscopically before and after the work extraction.

From the condition 1, by the work extraction, the state of  $E_W$  is changed from a microcanonical state of energy (or a state close to the microcanonical state) to another microcanonical state of energy (or a state close to the microcanonical state). Because of the condition 2, the entropy of the initial state of  $E_W$  (=a microcanonical state) is equal to the entropy of the final state of  $E_W$  (=another microcanonical state). Therefore, the density of states of  $E_W$  should be constant with energy. This condition means that  $E_W$  is not a thermodynamic system, but a mechanical system. (Fig. 1).

Mathematically, we require  $E_W$  to satisfy the following three conditions:

(A): The Hamiltonian  $\hat{H}_{E_W}^{(n)} := \sum_{e \in \Lambda^{(n)}} |e\rangle_{E_W} E_W \langle e|$  does not have degeneracy. The Hamiltonian  $\hat{H}_{E_W}^{(n)}$  has the ground state, i.e., there exists the minimum in  $\Lambda^{(n)}$ . We fix the energy eigenvalue of the ground state as 0.

(B): For arbitrary energy eigenvalues  $h$  and  $h'$  of  $H_I^{(n)}$  and an arbitrary  $e \in \Lambda^{(n)}$  satisfying  $e + h - h' > 0$ , there exists a  $e' \in \Lambda^{(n)}$  satisfying  $e' = e + h - h'$ .

(C): The density of states of  $E_W$  is constant with energy. Namely, for an arbitrary positive number  $\delta > 0$ , the dimension  $\Omega_{E_W}^{(n)}(U, \delta)$  of the following subspace of  $\mathcal{H}_{E_W}^{(n)}$  is

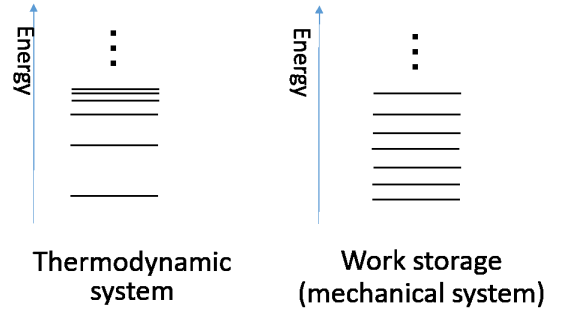


FIG. 1: Difference between thermodynamic systems and work storage. In the thermodynamic system, the energy levels become dense as the energy rises. In the work storage, the density of the energy levels is independent of energy.

independent of  $U$  whenever  $U - \delta > 0$  holds:

$$\mathcal{H}_{E_W}^{(n)}(U, \delta) = \text{span}\{|\psi\rangle | H_{E_W}^{(n)}|\psi\rangle = \lambda|\psi\rangle, \\ n(U - \delta) \leq \lambda \leq n(U + \delta)\} \quad (13)$$

*Dynamics: randomized energy preserving unitary* We next formulate the dynamics on  $IE_W$  as the randomized unitary dynamics  $\{p_j^{(n)}, V_j^{(n)}\}$ . Here,  $\{p_j^{(n)}\}$  is an arbitrary probability distribution, and  $\{V_j^{(n)}\}$  is a set of the unitary transformations on  $IE_W$  satisfying the following two features:

Feature 1 (Energy preserving): each unitary  $V_j^{(n)}$  preserves the energy of  $IE_W$ :

$$[V_j^{(n)}, H_I^{(n)} + H_{E_W}^{(n)}] = 0. \quad (14)$$

Feature 2 (Shift-invariant): each unitary  $V_j^{(n)}$  is stable for the (not so large) translation of the initial state of  $E_W$  in energy. Namely, we require that there will be no change in the dynamics of  $I$  when we perform the translation the initial state of  $E_W$  in energy. The mathematical definition is given as follows: Let us take positive numbers  $\tilde{U}_W$  and  $h_0$ . We say that “ $V_j^{(n)}$  is shift-invariant in  $h_0$ -neighbourhood of  $\tilde{U}_W$ ,” if and only if the following relation holds for arbitrary number  $h$  satisfying  $-h_0 < h < h_0$ , an arbitrary state  $\rho_I$  of  $I$ , and an arbitrary state  $\sigma_{E_W}$  of  $E_W$  whose support  $\text{supp}[\sigma_{E_W}]$  is on the subspace  $\mathcal{H}_{E_W}^{(n)}(\tilde{U}_W, h_0)$ :

$$\text{Tr}_{E_W}[V_j^{(n)}(\rho_I \otimes \sigma_{E_W})V_j^{(n)\dagger}] \\ = \text{Tr}_{E_W}[V_j^{(n)}(\rho_I \otimes \Gamma_h(\sigma_{E_W}))V_j^{(n)\dagger}], \quad (15)$$

where,  $\Gamma_h$  is the energy translation unitary of  $E_W$  [26]

$$\Gamma_h(\sigma_{E_W}) := \sum_e |e + h\rangle \langle e|_{\sigma_{E_W}} \sum_{e'} |e'\rangle \langle e' + h|. \quad (16)$$

Under the above formulation, we define the work extraction by the adiabatic transformation as follows:

*Adiabatic work extraction:* Let us fix arbitrary sets of the inverse temperatures and the control parameters  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m)$  and  $(\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$ , and fix an arbitrary real number  $\tilde{W}$  and an arbitrary positive number  $\delta > 0$ . When there exist real positive numbers  $\tilde{U}_W$  and  $h_0 > \delta$  and a randomized unitary  $\{p_j^{(n)}, V_j^{(n)}\}$  whose unitaries  $\{V_j^{(n)}\}$  are the energy-preserving and shift-invariant in  $h_0$ -neighbourhood of  $\tilde{U}_W$  and satisfy the following three conditions for sufficiently large  $n$ , we say that the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad}:\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is possible:

$$\lim_{n \rightarrow \infty} \|\sigma_I^{(n)} - \rho_{\tilde{\beta}'|\tilde{\lambda}'}^{(n)} \otimes |\tilde{\lambda}'\rangle\langle\tilde{\lambda}'|\| = 0, \quad (17)$$

$$\lim_{n \rightarrow \infty} \|\sigma_{E_W}^{(n)} - \pi_{\tilde{U}_W + \tilde{W}, \delta}^{(n)}\| = 0 \quad (18)$$

$$\lim_{n \rightarrow \infty} \frac{S(\sigma_{E_W}^{(n)}) - S(\pi_{\tilde{U}_W, \delta}^{(n)})}{n} = 0 \quad (19)$$

where  $\pi_{\tilde{U}, \delta}^{(n)}$  is the maximally mixed state of  $\mathcal{H}_{E_W}^{(n)}(\tilde{U}, \delta)$ ,  $\sigma_{IE_W}^{(n)} := \sum_j p_j^{(n)} V_j^{(n)} (\rho_I^{(n)} \otimes \pi_{\tilde{U}_W, \delta}^{(n)}) V_j^{(n)\dagger}$ ,  $\sigma_I^{(n)} := \text{Tr}_{E_W}[\sigma_{IE_W}^{(n)}]$  and  $\sigma_{E_W}^{(n)} := \text{Tr}_I[\sigma_{IE_W}^{(n)}]$ .

### III. MAIN RESULT

Under the above formulation, the following theorem holds:

**Theorem 1** *For arbitrary sets of the parameters  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m)$  and  $(\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  and an arbitrary real number  $\tilde{W}$ , the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad}:\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is possible if and only if the following two expressions hold:*

$$\Delta\tilde{S}_S \geq 0 \quad (20)$$

$$\tilde{W} = -\Delta\tilde{U}_S \quad (21)$$

where,  $\Delta\tilde{S}_S := \sum_l \tilde{S}_l(\beta'_l, \lambda'_l) - \sum_l \tilde{S}_l(\beta_l, \lambda_l)$  and  $\Delta\tilde{U}_S := \sum_l \tilde{U}_l(\beta'_l, \lambda'_l) - \sum_l \tilde{U}_l(\beta_l, \lambda_l)$ .

Theorem 1 is very simple, but it is different from the usual form of the first law and the second law of thermodynamics. Therefore, we derive the usual form of the first and the second law of thermodynamics for adiabatic processes as the corollaries of Theorem 1. Firstly, we derive the second law. The second law of thermodynamics focus on the convertibility from a combination of equilibrium states  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m)$  to another combination of equilibrium  $(\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  by the adiabatic transformation, without considering the amount of extracted work. Therefore, the statement that “the conversion  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m)$  to  $(\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  by the adiabatic transformation is possible” is equivalent to the statement “there exists a real number  $\tilde{W}$ , and the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad}:\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$ .” Hence, the second law is easily derived from Theorem 1:

**Corollary 1 (Second Law for adiabatic processes)** *The conversion  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m)$  to  $(\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  by the adiabatic transformation is possible if and only if  $\Delta\tilde{S}_S \geq 0$  holds.*

Similarly, the first law is easily derived from Theorem 1:

**Corollary 2 (First Law for adiabatic processes)** *When the conversion  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m)$  to  $(\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  by the adiabatic transformation is possible, the value of the extracted work  $\tilde{W}$  during the adiabatic transformation satisfies the following equality:*

$$\tilde{W} = -\Delta\tilde{U}_S. \quad (22)$$

We emphasize that Theorem 1 is derived from the assumption that the probability of energy of Gibbs states satisfies the large deviation condition (7) and (8). We do not impose stronger assumptions, including i.i.d. feature.

In thermodynamics, from the first law and the second law for adiabatic processes, the other forms of the first law and the second law are derived, e.g. the principle of maximal work and Carnot’s theorem and Clausius inequality. From Theorem 1, we can derive other forms of the first law and the second law of thermodynamics, by the same way as in thermodynamics. In the next subsection, we derive the principle of the first law and the second law for the isothermal processes as examples.

#### A. Corollary: First Law and Second Law of Thermodynamics for Isothermal Processes

Now, let us consider the isothermal processes. We fix an inverse temperature  $\beta$  arbitrarily. Let us take two thermodynamic systems  $S$  and  $B_M$ . The system  $S$  is the working body, and the system  $B_M$  is the heat bath. Here,  $M$  is a real positive number which expresses the “size” of  $B_M$ , and we assume that the specific heat of  $B_M$  is proportional to  $M$ , at least when the temperature of  $B_M$  is close to  $1/\beta$ . We firstly define the possibility of the work extraction by isothermal process based on the adiabatic work extraction as follows:

*isothermal work extraction:* Let us fix the thermodynamic system  $S$ , and fix its control parameters  $\lambda_S$  and  $\lambda'_S$  arbitrarily. We also fix an arbitrary real number  $\tilde{W}$ . Then, we define “the isothermal work extraction  $(\beta, \lambda_S) \rightarrow_{\text{is}:\tilde{W}} (\beta, \lambda'_S)$  is possible” as follows:

There exist two functions  $\beta'_M$  and  $\tilde{W}_M$  of  $M$  satisfying  $\lim_{M \rightarrow \infty} \beta'_M = \beta$  and  $\lim_{M \rightarrow \infty} \tilde{W}_M = \tilde{W}$ , and the adiabatic work extraction  $(\beta, \lambda_S) \times (\beta, \lambda_B) \rightarrow_{\text{ad}:\tilde{W}_M} (\beta, \lambda'_S) \times (\beta'_M, \lambda_B)$  is possible for sufficiently large  $M$ .

Under the above definition, the following two corollaries hold:

**Corollary 3 (Principle of Maximal Work)** *The isothermal work extraction  $(\beta, \lambda) \rightarrow_{\text{is}:\tilde{W}} (\beta, \lambda')$  is*

possible if and only if the following inequality holds:

$$\tilde{W} \leq -\Delta\tilde{F}_S \quad (23)$$

where

$$\Delta\tilde{F}_S := \tilde{F}_S(\beta, \lambda'_S) - \tilde{F}_S(\beta, \lambda_S). \quad (24)$$

#### Corollary 4 (First Law of Thermodynamics)

When the isothermal work extraction  $(\beta, \lambda) \rightarrow_{\text{is:}\tilde{W}} (\beta, \lambda')$  is possible, there exists at least one pair of  $\tilde{W}_M$  and  $\beta'_M$  satisfying the conditions in the definition of the isothermal work extraction. For any such pair, the limit  $\tilde{Q} := \lim_{M \rightarrow \infty} (\tilde{U}_{B_M}(\beta, \lambda_B) - \tilde{U}_{B_M}(\beta'_M, \lambda_B))$  exists, which is independent of  $\tilde{W}_M$  and  $\beta'_M$ , and satisfies

$$\tilde{W} = \tilde{Q} - \Delta\tilde{U}_S. \quad (25)$$

We prove these two corollaries in Appendix C.

#### IV. CONCLUSION

We have shown that the first law and the second law are derived from an assumption that “probability distributions of energy in Gibbs states satisfy large deviation”,

which is widely accepted as a property of thermodynamic equilibrium states. As the second law, we have shown that an adiabatic transformation from a set of Gibbs states to another set of Gibbs states is possible if and only if the regularized von Neumann entropy becomes large:

$$\Delta\tilde{S}_S \geq 0. \quad (26)$$

As the first law, we have shown that the energy loss of the thermodynamic system during the adiabatic transformation is stored in the work storage as “work,” in the following meaning; (i) the energy of the work storage takes certain values macroscopically, in the initial state and the final state. (ii) the entropy of the work storage in the final state is macroscopically equal to the entropy of the initial state.

As corollaries, our results gives other forms of the first law and the second law. As an example, we give the maximum work principle  $W \leq -\Delta F$  and the first law  $W = -\Delta U_S + Q$  in the isothermal processes.

*Acknowledgments:* The authors are grateful to Prof. Hiroshi Nagaoka and Mr. Naoto Shiraishi for helpful comments. This work was partially supported by JSPS Grant-in-Aid for Scientific Research (C) No. 16K00012 (TO).

- 
- [1] Boltzmann, L. E. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Wien Ber.* **66**, 275 (1872).
  - [2] Landau, L. D. and Lifshitz, E. M. *Statistical Physics. Vol. 5 of the Course of Theoretical Physics*, 3rd edition (1976).
  - [3] E. H. Lieb and J. Yngvason *Phys. Rep.* **310**, 1, (1996).
  - [4] A. Shimizu, *Basis of thermodynamics*, (University of Tokyo Press (in Japanese), ISBN: 4130626094, (2007).)
  - [5] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690, (1997).
  - [6] H. Tasaki, arXiv:cond-mat/0009244 (2000).
  - [7] J. Kurchan, arXiv:cond-mat/0007360(2000).
  - [8] S. DeLiberato and M. Ueda, *Phys. Rev. E* **84**, 051122 (2011).
  - [9] G. Xiao and J. Gong, arXiv:1503.00784, (2015).
  - [10] H. Tasaki, arXiv:1511.01999 (2015).
  - [11] A. Lenard, *J. Stat. Phys.*, **19**, 6, 575(1978).
  - [12] G. E. Crooks, *Phys. Rev. E* **60**, 2721 (1999).
  - [13] K. Sekimoto, *Stochastic Energetics (Lecture Notes in Physics)*, Springer, (2010).
  - [14] M. F. Frenzel, D. Jennings, and T. Rudolph, *Phys. Rev. E* **90**, 052136 (2014)
  - [15] M. Horodecki and J. Oppenheim, *Nat. Commun.* **4**, 2059 (2013).
  - [16] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, *New. J. Phys.* **13**, 053015, (2011).
  - [17] J. Aberg, *Nat. Commun.* **4**, 1925 (2013).
  - [18] D. Egloff, O. C. O. Dahlsten, R. Renner and V. Vedral, arXiv:1207.0434, (2012).
  - [19] F. G. S. L. Brandao, M. Horodecki, N. H. Y. Ng, J. Oppenheim, and S. Wehner, *PNAS*, 112,3215(2015).
  - [20] S. Popescu, arXiv:1009.2536.(2010).
  - [21] P. Skrzypczyk, A. J. Short and S. Popescu, *Nature Communications* **5**, 4185, (2014).
  - [22] Y. Guryanova, S. Popescu, A. J. Short, R. Silva and P. Skrzypczyk, *Nat. Comm.* **7**, 12049, (2016).
  - [23] L. Rio, J. Aberg, R. Renner, O. Dahlsten, and V. Vedral, *Nature* **474**, 61, (2011).
  - [24] M. Weilenmann, L. Krämer, P. Faist, and R. Renner, arXiv:1501.06920(2015).
  - [25] G. Gour, M. P. Muller, V. Narasimhachar, R. W. Spekkens, N. Y. Halpern, arXiv:1309.6586 (2013).
  - [26] J. Åberg, *Phys. Rev. Lett.* **113**, 150402 (2014).
  - [27] A. S. L. Malabarba, A. J. Short, P. Kammerlander, *New. J. Phys.* **17**, 045027 (2015).
  - [28] H. Tajima and M. Hayashi, “Optimal efficiency with finite-size heat baths,” arXiv: 1405.6457 (2014).
  - [29] M. Hayashi and H. Tajima, “Measurement-based Formulation of Quantum Heat Engine,” arXiv:1504.06150, (2015).
  - [30] H. Tajima and E. Wakakuwa, “Regularized Boltzmann entropy determines macroscopic adiabatic accessibility,” arXiv:1601.00487, (2016).
  - [31] H. Touchette, *Physics Reports* **478**, 1 (2009).
  - [32] R. S. Ellis, *Entropy, Large Deviations and Statistical Mechanics*, Springer, 1985
  - [33] T. S. Han: *Information-Spectrum Methods in Information Theory*, (Springer, Berlin Heidelberg New York, 2002) (originally appeared in Japanese in 1998)
  - [34] Y. Jiao, E. Wakakuwa and T. Ogawa, “Asymptotic Convertibility of Entanglement: A General Approach to Entanglement Concentration and Di-

lution,” AQIS2016 conference proceedings (2016).

http://aqis-conf.org/2016/wp-content/uploads/2015/12/0830-Tuesday-Page-2-Only.pdf

[35] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).

[36] M. A. Nielsen and I. L. Chuang, *Quantum Computation*

and Quantum Information (Cambridge University Press, Cambridge, 2000).

## Appendix A: Proof of Theorem 1

Hereafter, we use the abbreviations  $\tilde{U}_S(\vec{\beta}, \vec{\lambda}) := \sum_k \tilde{U}_k(\beta_k, \lambda_k)$ ,  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) := \sum_k \tilde{S}_k(\beta_k, \lambda_k)$ ,  $\beta_{\max} := \max\{\beta_1, \dots, \beta_m\}$  and  $\beta''_{\max} := \max\{\beta_1, \dots, \beta_m, \beta'_1, \dots, \beta'_m\}$ .

We prove Theorem 1 with using the following lemma:

**Lemma 1** *Let us assume that  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) < \tilde{S}_S(\vec{\beta}', \vec{\lambda}')$  holds. Let us fix an arbitrary positive real number  $\delta > 0$ . We also fix an arbitrary real positive numbers  $\tilde{U}_W$  and  $h_0$  satisfying*

$$h_0 > \delta \quad (A1)$$

$$\tilde{U}_W > |\tilde{U}_S(\vec{\beta}, \vec{\lambda}) - \tilde{U}_S(\vec{\beta}', \vec{\lambda}')| + 8h_0. \quad (A2)$$

Then, for an arbitrary  $\delta' > 0$  satisfying  $\delta' < \tilde{S}_S(\vec{\beta}', \vec{\lambda}') - \tilde{S}_S(\vec{\beta}, \vec{\lambda})$  and  $\delta' < 2\beta''_{\max}\delta$ , there exist a natural number  $N_{\delta'}$  and randomized unitary  $\{q_{j,\delta'}^{(n)}, V_{j,\delta'}^{(n)}\}_{n=N_{\delta'}}$  whose unitaries  $\{V_j^{(n)}\}$  are the energy-preserving and shift-invariant in  $h_0$ -neighbourhood of  $\tilde{U}_W$  and satisfy the following three conditions for arbitrary  $n \geq N_{\delta'}$ :

$$\|\sigma_{I,\delta'}^{(n)} - \rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| < \delta' \quad (A3)$$

$$\|\sigma_{E_W,\delta'}^{(n)} - \pi_{\tilde{U}_W - \Delta\tilde{U}_S,\delta}^{(n)}\| < \left(1 + \frac{4}{\beta''_{\max}\delta}\right) \delta' \quad (A4)$$

$$\frac{|S(\sigma_{E_W,\delta'}^{(n)}) - S(\pi_{\tilde{U}_W,\delta}^{(n)})|}{n} < \delta', \quad (A5)$$

where

$$\sigma_{IE_W,\delta'}^{(n)} := \sum_j q_{j,\delta'}^{(n)} V_{j,\delta'}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}| \otimes \pi_{\tilde{U}_W,\delta}^{(n)} V_{j,\delta'}^{(n)\dagger}, \quad (A6)$$

and  $\sigma_{I,\delta'}^{(n)} := \text{Tr}_{E_W}[\sigma_{IE_W,\delta'}^{(n)}]$  and  $\sigma_{E_W,\delta'}^{(n)} := \text{Tr}_I[\sigma_{IE_W,\delta'}^{(n)}]$ .

Lemma 1 is proved in Appendix C.

Let us prove Theorem 1. We firstly show that when  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) < \tilde{S}_S(\vec{\beta}', \vec{\lambda}')$  and  $\tilde{W} = -\Delta\tilde{U}_S$  hold, the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad}:\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is possible. Let us fix an arbitrary  $\delta > 0$ . We also take the positive numbers  $h_0$  and  $\tilde{U}_W$  satisfying (A1) and (A2). We refer to the smallest natural number  $k$  satisfying  $\frac{1}{k} \leq \tilde{S}_S(\vec{\beta}', \vec{\lambda}') - \tilde{S}_S(\vec{\beta}, \vec{\lambda})$  and  $\frac{1}{k} \leq 2\beta''_{\max}\delta$  as  $k_0$ . For a sequence of natural numbers  $\{k\}_{k=k_0}^\infty$ , we define  $\delta'_k := \frac{1}{k}$  and take the natural number  $N_{\delta'_k}$  and the randomized unitary  $\{q_{j,\delta'_k}^{(n)}, V_{j,\delta'_k}^{(n)}\}_{n=N_{\delta'_k}}$  by substituting  $\delta'_k$  for  $\delta'$  in Lemma 1. We also define a natural number  $\{\tilde{N}_k\}_{k=k_0}^\infty$  as

$$\begin{aligned} \tilde{N}_{k_0} &:= \tilde{N}_{\delta'_{k_0}}, \\ \tilde{N}_k &:= \max\{\tilde{N}_{\delta'_k}, \tilde{N}_{\delta'_{k-1}} + 1\}. \end{aligned} \quad (A7)$$

We also define a randomized shift-invariant and energy preserving unitary  $\{q_j^{(n)}, V_j^{(n)}\}_{n=\tilde{N}_{k_0}}^\infty$  as

$$q_j^{(n)} := q_{j,\delta'_k}^{(n)} \quad (\text{for } \tilde{N}_{k+1} - 1 \geq n \geq \tilde{N}_k) \quad (A8)$$

$$V_j^{(n)} := V_{j,\delta'_k}^{(n)} \quad (\text{for } \tilde{N}_{k+1} - 1 \geq n \geq \tilde{N}_k). \quad (A9)$$

Then,  $\{q_j^{(n)}, V_j^{(n)}\}_{n=\tilde{N}_{k_0}}^\infty$  satisfies (17)–(19) because of (A3)–(A5) and  $\tilde{W} = -\Delta\tilde{U}_S$ . Therefore, when  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) < \tilde{S}_S(\vec{\beta}', \vec{\lambda}')$  and  $\tilde{W} = -\Delta\tilde{U}_S$  hold, the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad}:\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is possible.

We secondly show that when  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) = \tilde{S}_S(\vec{\beta}', \vec{\lambda}')$  and  $\tilde{W} = -\Delta\tilde{U}_S$  hold, the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad}:\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is possible. We refer to the smallest natural number  $t$  satisfying  $\frac{1}{t} < 2\beta''_{\max}\delta$  as  $t_0$ , and define a sequence of natural numbers  $\{t\}_{t=t_0}^\infty$ . For each natural number  $t$ , we define  $\vec{\beta}'_t := (\beta'_1, \dots, \beta'_{m-1}, \beta'_m(1 - \frac{1}{t}))$ . Then, because of  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) < \tilde{S}_S(\vec{\beta}', \vec{\lambda}')$  and Lemma 1, there exist a natural number  $N''_t$  and a randomized unitary  $\{q_{j,t}^{(n)}, V_{j,t}^{(n)}\}_{n=N''_t}^\infty$  satisfying the followings:

$$\|\sigma_{I,t}^{(n)} - \rho_{\vec{\beta}'_t|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| < \frac{1}{t} \quad (\text{A10})$$

$$\|\sigma_{EW,t}^{(n)} - \pi_{\tilde{U}_W - \Delta_t \tilde{U}_S, \delta}^{(n)}\| < \left(1 + \frac{4}{\beta''_{\max}\delta}\right) \frac{1}{t} \quad (\text{A11})$$

$$\frac{|S(\sigma_{EW,t}^{(n)}) - S(\pi_{\tilde{U}_W, \delta}^{(n)})|}{n} < \frac{1}{t}. \quad (\text{A12})$$

where  $\Delta_t \tilde{U}_S := \tilde{U}_S(\vec{\beta}'_t, \vec{\lambda}') - \tilde{U}_S(\vec{\beta}, \vec{\lambda})$  and

$$\sigma_{IEW,t}^{(n)} := \sum_j q_{j,t}^{(n)} V_{j,t}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}| \otimes \pi_{\tilde{U}_W, \delta}^{(n)} V_{j,t}^{(n)\dagger}, \quad (\text{A13})$$

$$\sigma_{I,t}^{(n)} := \text{Tr}_{EW}[\sigma_{IEW,t}^{(n)}], \quad \sigma_{EW,t}^{(n)} := \text{Tr}_I[\sigma_{IEW,t}^{(n)}]. \quad (\text{A14})$$

We can prove the above by substituting  $1/t$  for  $\delta'$  in Lemma 1. Now, we define

$$f(t) := \|\rho_{\vec{\beta}'_t|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'| - \rho_{\vec{\beta}, \vec{\lambda}}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| \quad (\text{A15})$$

$$g(t) := \pi_{\tilde{U}_W - \Delta_t \tilde{U}_S, \delta}^{(n)}. \quad (\text{A16})$$

Because  $\tilde{U}$  and  $\tilde{S}$  are continuous functions of  $\beta$ , we obtain  $\lim_{t \rightarrow \infty} f(t) = \lim_{t \rightarrow \infty} g(t) = 0$  and

$$\|\sigma_{I,t}^{(n)} - \rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| < \frac{1}{t} + f(t) \quad (\text{A17})$$

$$\|\sigma_{EW,t}^{(n)} - \pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)}\| < \left(1 + \frac{4}{\beta''_{\max}\delta}\right) \frac{1}{t} + g(t) \quad (\text{A18})$$

$$\frac{|S(\sigma_{EW,t}^{(n)}) - S(\pi_{\tilde{U}_W, \delta}^{(n)})|}{n} < \frac{1}{t}. \quad (\text{A19})$$

Therefore, we define  $\{\tilde{N}''_t\}_{t=t_0}^\infty$  as

$$\tilde{N}''_{t_0} = N''_{t_0} \quad (\text{A20})$$

$$\tilde{N}''_t = \max\{N''_t, N''_{t-1} + 1\}, \quad (\text{A21})$$

and define  $\{\tilde{q}_{j,t}^{(n)}, \tilde{V}_{j,t}^{(n)}\}_{n=N''_{t_0}}^\infty$  as

$$\tilde{q}_j^{(n)} := q_{j,t}^{(n)} \quad (\text{for } \tilde{N}''_{t+1} - 1 \geq n \geq \tilde{N}''_t) \quad (\text{A22})$$

$$\tilde{V}_j^{(n)} := V_{j,t}^{(n)} \quad (\text{for } \tilde{N}''_{t+1} - 1 \geq n \geq \tilde{N}''_t). \quad (\text{A23})$$

Then,  $\{\tilde{q}_{j,t}^{(n)}, \tilde{V}_{j,t}^{(n)}\}_{n=N''_{t_0}}^\infty$  satisfies (17)–(19) because of (A17)–(A19) and  $\tilde{W} = \Delta\tilde{U}_S$ .

Next, we show that when  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) > \tilde{S}_S(\vec{\beta}', \vec{\lambda}')$  or  $\tilde{W} \neq -\Delta\tilde{U}_S$  holds, the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad}:\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is impossible. Let us assume that  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad}:\tilde{W}}$

$(\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  would be possible even when  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) > \tilde{S}_S(\vec{\beta}', \vec{\lambda}')$  holds. Then, there exists  $\{p_j^{(n)}, V_j^{(n)}\}$  satisfying (17)–(19). For the simplicity in expressions, hereafter we use the following description:

$$\Lambda_I^{\rho_{EW}}(\rho_I) := \text{Tr}_{EW}[\sum_j p_j^{(n)} V_j^{(n)} \rho_I \otimes \rho_{EW} V_j^{(n)\dagger}], \quad (\text{A24})$$

$$\Lambda_{EW}^{\rho_I}(\rho_{EW}) := \text{Tr}_I[\sum_j p_j^{(n)} V_j^{(n)} \rho_I \otimes \rho_{EW} V_j^{(n)\dagger}]. \quad (\text{A25})$$

where  $\rho_I$  and  $\rho_{EW}$  are arbitrary states of  $I$  and  $EW$ , respectively.

Let us derive a contradiction from the existence of  $\{p_j^{(n)}, V_j^{(n)}\}$ . Firstly, we define the region  $1_{Con}$  as the subspace of  $\mathcal{H}_I^{(n)}$  which is spanned by the energy eigenvectors of  $H_I^{(n)}$  whose eigenvalues  $E^{(n)}$  satisfy

$$n(\tilde{U}_S(\vec{\beta}, \vec{\lambda}) - \epsilon) \leq E^{(n)} \leq n(\tilde{U}_S(\vec{\beta}, \vec{\lambda}) + \epsilon) \quad (\text{A26})$$

where “Con” is abbreviation of “Converse part.” Next, we take the following approximate state of  $\rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|$ :

$$\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)Con} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}| := \frac{\Pi_{\text{region}1_{Con}}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}| \Pi_{\text{region}1_{Con}}^{(n)}}{\text{Tr}[\Pi_{\text{region}1_{Con}}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|]} \quad (\text{A27})$$

where  $\Pi_{\text{region}1_{Con}}^{(n)}$  is the projection to the region  $1_{Con}$ .

Next, we refer to the support of  $\Lambda_I^{\pi_{\tilde{U}_W, \delta}}(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)Con})$  as the region  $2'_{Con}$ . We also define the region  $2_{Con}$  as the subspace of  $\mathcal{H}_I^{(n)}$  which is spanned by the energy eigenvectors of  $H_I^{(n)}$  whose eigenvalues  $E^{(n)}$  satisfy

$$n(\tilde{U}_S(\vec{\beta}', \vec{\lambda}') - \epsilon') \leq E^{(n)} \leq n(\tilde{U}_S(\vec{\beta}', \vec{\lambda}') + \epsilon'). \quad (\text{A28})$$

Because the Hamiltonian  $H_{EW}^{(n)}$  has a ground energy level and each  $V_j^{(n)}$  is energy-preserving, if we take  $\epsilon' > 0$  enough large, the region  $2_{Con}$  includes the region  $2'_{Con}$ . Then, we define an approximate state of  $\rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|$  as follows:

$$\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)Con} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'| := \frac{\Pi_{\text{region}2_{Con}}^{(n)} \rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'| \Pi_{\text{region}2_{Con}}^{(n)}}{\text{Tr}[\Pi_{\text{region}2_{Con}}^{(n)} \rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|]} \quad (\text{A29})$$

where  $\Pi_{\text{region}2_{Con}}^{(n)}$  is the projection to the region  $2_{Con}$ .

Because of the large deviation assumptions (7) and (8), there exists a real positive number  $\alpha$ , the inequalities

$$\text{Tr}[\Pi_{\text{region}1_{Con}}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|] \geq 1 - e^{-n\alpha} \quad (\text{A30})$$

$$\text{Tr}[\Pi_{\text{region}2_{Con}}^{(n)} \rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|] \geq 1 - e^{-n\alpha} \quad (\text{A31})$$

hold for sufficiently large  $n$ . Therefore, because of the gentle measurement lemma [35], the following inequalities hold for the sufficiently large  $n$ :

$$\|\rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}| - \tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)Con} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|\| \leq e^{-n\alpha/2} \quad (\text{A32})$$

$$\|\rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'| - \tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)Con} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| \leq e^{-n\alpha/2} \quad (\text{A33})$$

Also, for arbitrary  $\epsilon > 0$ ,  $\vec{\beta}$  and  $\vec{\lambda}$ , the following relations hold:

$$\lim_{n \rightarrow \infty} \frac{S(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)Con} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|)}{n} = \tilde{S}_S(\vec{\beta}, \vec{\lambda}), \quad (\text{A34})$$

$$\lim_{n \rightarrow \infty} \frac{(1 - r_{\text{main}}^{Con}) S(\rho_{\vec{\beta}|\vec{\lambda}}^{(n)Con\cap} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|)}{n} = 0, \quad (\text{A35})$$



where

$$r_{\text{main}}^{\text{Con}} := \text{Tr}[\Pi_{\text{region1Con}}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|] \quad (\text{A36})$$

$$\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}\neg} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}| := \frac{(1 - \Pi_{\text{region1Con}}^{(n)}) \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}| (1 - \Pi_{\text{region1Con}}^{(n)})}{\text{Tr}[(1 - \Pi_{\text{region1Con}}^{(n)}) \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|]} \quad (\text{A37})$$

(Proof of (A34) and (A35): Note that the dimension of region 1<sub>Con</sub> is lower than  $e^{n(\beta_{\max}\tilde{U}_S(\vec{\beta},\vec{\lambda}) - \sum_k \tilde{F}_k(\beta_k, \lambda_k) + \alpha)}$  for sufficiently large  $n$ , where  $\beta_{\max} := \max\{\beta_1, \dots, \beta_m\}$ . The relations (A34) and (A35) are easily derived from the following inequality:

$$\begin{aligned} S(\rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|) &= S(\rho_{\vec{\beta}|\vec{\lambda}}^{(n)}) = S(r_{\text{main}}^{\text{Con}} \tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} + (1 - r_{\text{main}}^{\text{Con}}) \rho_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}\neg}) \\ &\geq r_{\text{main}}^{\text{Con}} S(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}}) + (1 - r_{\text{main}}^{\text{Con}}) S(\rho_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}\neg}) \\ &\stackrel{(b)}{\geq} r_{\text{main}}^{\text{Con}} (S(\rho_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}}) - e^{-n\alpha/2} \log e^{n(\beta_{\max}\tilde{U}_S(\vec{\beta},\vec{\lambda}) - \sum_k \tilde{F}_k(\beta_k, \lambda_k) + \alpha)}) + (1 - r_{\text{main}}^{\text{Con}}) S(\rho_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}\neg}) \end{aligned} \quad (\text{A38})$$

where (b) is shown by Fannes inequality, (A31) and (A32).

Because of (17), we obtain

$$\lim_{n \rightarrow \infty} \|\Lambda_I^{\pi_{\tilde{U}_W, \delta}^{(n)}}(\rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|) - \rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| = 0. \quad (\text{A39})$$

Because of (A32), we obtain

$$\lim_{n \rightarrow \infty} \|\Lambda_I^{\pi_{\tilde{U}_W, \delta}^{(n)}}(\rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|) - \Lambda_I^{\pi_{\tilde{U}_W, \delta}^{(n)}}(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|)\| = 0. \quad (\text{A40})$$

Because of (A33), we obtain

$$\lim_{n \rightarrow \infty} \|\rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'| - \tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)\text{Con}} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| = 0. \quad (\text{A41})$$

Therefore,

$$\lim_{n \rightarrow \infty} \|\Lambda_I^{\pi_{\tilde{U}_W, \delta}^{(n)}}(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|) - \tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)\text{Con}} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| = 0. \quad (\text{A42})$$

Because of (A28), for sufficiently large natural number  $D$ , the dimension of the region 2<sub>Con</sub> is lower than  $D^n$ . Therefore, we obtain the following from the Fannes inequality:

$$\lim_{n \rightarrow \infty} \frac{1}{n} (S(\Lambda_I^{\pi_{\tilde{U}_W, \delta}^{(n)}}(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|)) - S(\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)\text{Con}} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|)) = 0 \quad (\text{A43})$$

Therefore, because of (A34),

$$\lim_{n \rightarrow \infty} \frac{1}{n} S(\Lambda_I^{\pi_{\tilde{U}_W, \delta}^{(n)}}(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|)) = \tilde{S}_S(\vec{\beta}', \vec{\lambda}'). \quad (\text{A44})$$

Note that  $\Lambda_{E_W}^{\rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|}(\pi_{\tilde{U}_W, \delta}) = r_{\text{main}}^{\text{Con}} \Lambda_{E_W}^{\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|}(\pi_{\tilde{U}_W, \delta}) + (1 - r_{\text{main}}^{\text{Con}}) \Lambda_{E_W}^{\tilde{\rho}_{\vec{\beta}}^{(n)-\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|}(\pi_{\tilde{U}_W, \delta})$ . Therefore,

$$\begin{aligned} S(\Lambda_{E_W}^{\rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|}(\pi_{\tilde{U}_W, \delta})) &\geq r_{\text{main}}^{\text{Con}} S(\Lambda_{E_W}^{\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|}(\pi_{\tilde{U}_W, \delta})) + (1 - r_{\text{main}}^{\text{Con}}) S(\Lambda_{E_W}^{\tilde{\rho}_{\vec{\beta}}^{(n)-\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|}(\pi_{\tilde{U}_W, \delta})) \\ &\geq r_{\text{main}}^{\text{Con}} S(\Lambda_{E_W}^{\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|}(\pi_{\tilde{U}_W, \delta})). \end{aligned} \quad (\text{A45})$$

Note that for  $\rho'_A := \text{Tr}_B[U_{AB} \rho_A \otimes \rho_B U_{AB}^\dagger]$  and  $\rho'_B := \text{Tr}_A[U_{AB} \rho_A \otimes \rho_B U_{AB}^\dagger]$ ,  $S(\rho_A) + S(\rho_B) \leq S(\rho'_A) + S(\rho'_B)$  holds. Therefore, because of (19) and (A45), we obtain

$$\liminf_{n \rightarrow \infty} \frac{1}{n} (S(\Lambda_I^{\pi_{\tilde{U}_W, \delta}^{(n)}}(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|)) - S(\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)\text{Con}} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|)) \geq 0. \quad (\text{A46})$$

Because of (A34),  $S(\tilde{\rho}_{\tilde{\beta}|\tilde{\lambda}}^{(n)Con} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}|)/n$  goes to  $\tilde{S}_S(\tilde{\beta}, \tilde{\lambda})$  at the limit of  $n \rightarrow \infty$ . We have assumed that  $\tilde{S}_S(\tilde{\beta}, \tilde{\lambda}) > \tilde{S}_S(\tilde{\beta}', \tilde{\lambda}')$ . Therefore (A46) implies that  $S(\Lambda_I^{\pi_{\tilde{U}_W, \delta}^{(n)}}(\tilde{\rho}_{\tilde{\beta}|\tilde{\lambda}}^{(n)Con} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}|))/n$  cannot converge to  $\tilde{S}_S(\tilde{\beta}', \tilde{\lambda}')$  at the limit of  $n \rightarrow \infty$ . This contradicts to (A44). Therefore, when  $\tilde{S}_S(\tilde{\beta}, \tilde{\lambda}) > \tilde{S}_S(\tilde{\beta}', \tilde{\lambda}')$  holds, the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad:}\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is impossible.

Next, we show that when  $\tilde{S}_S(\tilde{\beta}, \tilde{\lambda}) \leq \tilde{S}_S(\tilde{\beta}', \tilde{\lambda}')$  and  $\tilde{W} \neq -\Delta\tilde{U}_S$  hold, the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad:}\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is impossible. Let us assume that  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad:}\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  would be possible even when  $\tilde{S}_S(\tilde{\beta}, \tilde{\lambda}) \leq \tilde{S}_S(\tilde{\beta}', \tilde{\lambda}')$  and  $\tilde{W} \neq -\Delta\tilde{U}_S$  hold. Let us take a real number  $\delta$  such that  $0 < \delta < |-\Delta\tilde{U}_S - \tilde{W}|/10$ . Then, there exist  $\tilde{U}_W$ ,  $h_0$  and  $\{p_j^{(n)}, V_j^{(n)}\}$  satisfying (17)–(19). Let us derive a contradiction from the existence of  $\{p_j^{(n)}, V_j^{(n)}\}$ .

Because of (17), the following holds for sufficiently large  $n$ :

$$\|\sigma_I^{(n)} - \rho_{\tilde{\beta}'|\tilde{\lambda}'}^{(n)} \otimes |\tilde{\lambda}'\rangle\langle\tilde{\lambda}'|\| < \frac{1}{100} \quad (\text{A47})$$

Also, (7) and (8) imply the following holds for sufficiently large  $n$ :

$$\text{Tr}[\Pi_{\tilde{U}_S(\tilde{\beta}', \tilde{\lambda}'), \delta}^{(n)I} \rho_{\tilde{\beta}'|\tilde{\lambda}'}^{(n)} \otimes |\tilde{\lambda}'\rangle\langle\tilde{\lambda}'|] > \frac{99}{100}, \quad (\text{A48})$$

where  $\Pi_{X,Y}^{(n)I}$  is the proposition to the subspace of  $\mathcal{H}_I^{(n)}$  spanned by the energy eigenvectors whose eigenvalues  $E^{(n)}$  satisfy

$$n(X - Y) \leq E^{(n)} \leq n(X + Y). \quad (\text{A49})$$

Hereafter, we also use  $\Pi_{X,Y}^{(n)E_W}$  and  $\Pi_{X,Y}^{(n)IE_W}$  as the propositions to the subspaces of  $\mathcal{H}_{E_W}^{(n)}$  and  $\mathcal{H}_{IE_W}^{(n)}$  spanned by the energy eigenvectors whose eigenvalues  $E^{(n)}$  satisfy (A49), respectively. From (A47) and (A48), we obtain

$$\text{Tr}[\Pi_{\tilde{U}_S(\tilde{\beta}', \tilde{\lambda}'), \delta}^{(n)I} \sigma_I^{(n)}] > \frac{98}{100}. \quad (\text{A50})$$

The large deviation assumptions (7) and (8) imply the following holds for sufficiently large  $n$ :

$$\text{Tr}[\Pi_{\tilde{U}_S(\tilde{\beta}, \tilde{\lambda}), \delta}^{(n)I} \rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}|] > \frac{99}{100}. \quad (\text{A51})$$

Therefore, for sufficiently large  $n$ ,

$$\text{Tr}[\Pi_{\tilde{U}_S(\tilde{\beta}, \tilde{\lambda}) + \tilde{U}_W, 2\delta}^{(n)IE_W} \rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}| \otimes \pi_{\tilde{U}_W, \delta}^{(n)}] > \frac{99}{100}. \quad (\text{A52})$$

Because each  $V_j^{(n)}$  is an energy preserving unitary, (A52) implies the following:

$$\text{Tr}[\Pi_{\tilde{U}_S(\tilde{\beta}, \tilde{\lambda}) + \tilde{U}_W, 2\delta}^{(n)IE_W} \sigma_{IE_W}^{(n)}] > \frac{99}{100}. \quad (\text{A53})$$

Because of (A50) and (A53), we obtain

$$\text{Tr}[\Pi_{-\Delta\tilde{U}_S + \tilde{U}_W, 4\delta}^{(n)E_W} \sigma_{E_W}^{(n)}] > \frac{98}{100} \quad (\text{A54})$$

for sufficiently large  $n$ . However, because of  $\delta < |-\Delta\tilde{U}_S - \tilde{W}|/10$ , the following also holds:

$$\text{Tr}[\Pi_{-\Delta\tilde{U}_S + \tilde{U}_W, 4\delta}^{(n)E_W} \pi_{\tilde{U}_W + \tilde{W}, \delta}^{(n)}] = 0. \quad (\text{A55})$$

(A54) and (A55) contradicts to (18). Therefore, when  $\tilde{S}_S(\tilde{\beta}, \tilde{\lambda}) \leq \tilde{S}_S(\tilde{\beta}', \tilde{\lambda}')$  and  $\tilde{W} \neq -\Delta\tilde{U}_S$  hold, the adiabatic work extraction  $(\beta_1, \lambda_1) \times \dots \times (\beta_m, \lambda_m) \rightarrow_{\text{ad:}\tilde{W}} (\beta'_1, \lambda'_1) \times \dots \times (\beta'_m, \lambda'_m)$  is impossible.

## Appendix B: Proof of Lemma 1

We prove Lemma 1 with using the following lemma:

**Lemma 2** Let  $\hat{\rho} = \{\rho_n\}_{n=1}^\infty$  and  $\hat{\sigma} = \{\sigma_n\}_{n=1}^\infty$  be general sequences of quantum states on  $\mathcal{H}_n$  whose dimension is  $d^n$  ( $n = 1, 2, \dots$ ). If  $\overline{H}(\hat{\rho}) < \underline{H}(\hat{\sigma})$  holds, then there exists a sequence of unital CPTP maps  $\mathcal{R}_n$  ( $n = 1, 2, \dots$ ) such that

$$\lim_{n \rightarrow \infty} \|\mathcal{R}_n(\rho_n) - \sigma_n\|_1 = 0. \quad (\text{B1})$$

Here,  $\overline{H}(\hat{\rho})$  and  $\underline{H}(\hat{\sigma})$  are sup- and inf- spectral entropy rates defined by

$$\begin{aligned} \overline{H}(\hat{\rho}) &:= \inf \left\{ \gamma : \liminf_{n \rightarrow \infty} \text{Tr} [\{\rho_n \geq e^{-n\gamma} I_n\} \rho_n] = 1 \right\}, \\ \underline{H}(\hat{\sigma}) &:= \sup \left\{ \gamma : \limsup_{n \rightarrow \infty} \text{Tr} [\{\sigma_n \geq e^{-n\gamma} I_n\} \sigma_n] = 0 \right\}, \end{aligned}$$

$I_n$  is the identity operator on  $\mathcal{H}_n$ , and  $\{\rho_n \geq 2^{-n\gamma} I_n\}$  is the projection operator on a subspace of  $\mathcal{H}_n$  spanned by the eigenvectors of  $\rho_n$  with eigenvalues not smaller than  $2^{-n\gamma}$ .

Moreover, when  $\rho_n$  and  $\sigma_n$  are diagonalized by common basis  $\{|x\rangle\}_{x=1}^{d^n}$ , there exist a probability  $\{q_j^{(n)}\}_{j=1}^{d^n}$  and a set of invertible functions  $\{f_j^{(n)}\}_{j=1}^{d^n}$  from  $\{x\}_{x=1}^{d^n}$  to  $\{x\}_{x=1}^{d^n}$  satisfying

$$\mathcal{R}_n(\rho_n) = \sum_{j=1}^{d^n} q_j^{(n)} |f_j^{(n)}(x)\rangle \langle x| \rho_n |x\rangle \langle f_j^{(n)}(x)|. \quad (\text{B2})$$

Lemma 2 has been given in [34].

When the sup and inf spectral entropies of  $\hat{\rho}$  satisfies  $\overline{H}(\hat{\rho}) = \underline{H}(\hat{\rho}) = H$  for a real number  $H$ , we say that  $\{\rho_n\}_{n=1}^\infty$  has the entropy spectrum  $H$ . Note that because of the large deviation assumptions (7) and (8), the sequence  $\{\rho_{\vec{\beta}|\vec{\lambda}}^{(n)}\}$  has the entropy spectrum  $\tilde{S}_S(\vec{\beta}, \vec{\lambda})$ .

**Proof of Lemma 1:** We construct the randomized unitary  $\{q_{j,\delta'}^{(n)}, V_{j,\delta'}^{(n)}\}$  satysfying (A3)–(A5) concretely. Firstly, we label the energy eigenstates of  $H_{S|\vec{\lambda}}^{(n)}$  as

$$|\vec{i}_{|\vec{\lambda}}\rangle := |(i_1)_{|\lambda_1}\rangle \otimes |(i_2)_{|\lambda_2}\rangle \otimes \dots \otimes |(i_m)_{|\lambda_m}\rangle, \quad (\text{B3})$$

where  $|(i_k)_{|\lambda_k}\rangle$  is an energy eigenstate of  $H_{S_k|\lambda_k}^{(n)}$  whose engenvalue is the  $i_k$ -th smallest eigenvalue of  $H_{k|\lambda_k}^{(n)}$ .

Secondly, we define the following subspace of  $\mathcal{H}_S^{(n)}$  as a function of  $\vec{\beta}$ ,  $\vec{\lambda}$  and a real positive number  $\epsilon$ :  
*Region- $[\vec{\beta}, \vec{\lambda}, \epsilon]$ :* The subspace of  $\mathcal{H}_S^{(n)}$  spanned by  $|\vec{i}_{|\vec{\lambda}}\rangle$  whose energy eigenvalues  $E_{k|\lambda_k}^{(n)}(i_k)$  of  $H_{k|\lambda_k}^{(n)}$  satisfying

$$n \left( \tilde{U}_k(\beta_k, \lambda_k) - \frac{\epsilon}{m} \right) \leq E_{k|\lambda_k}^{(n)}(i_k) \leq n \left( \tilde{U}_k(\beta_k, \lambda_k) + \frac{\epsilon}{m} \right). \quad (\text{B4})$$

We refer to the projection to the region- $[\vec{\beta}, \vec{\lambda}, \epsilon]$  and the dimension of the region- $[\vec{\beta}, \vec{\lambda}, \epsilon]$  as  $\Pi_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}$  and  $D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}$ . Then, there exists a real positive number  $\alpha$ , the projection and dimension  $\Pi_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}$  and  $D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}$  satisfy the following relations for sufficiently large  $n$ :

$$\text{Tr}[\Pi_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)}] \geq 1 - e^{-n\alpha}, \quad (\text{B5})$$

$$e^{n(\tilde{S}_S(\vec{\beta}, \vec{\lambda}) - 2\beta_{\max}\epsilon)} \leq D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} \leq e^{n(\tilde{S}_S(\vec{\beta}, \vec{\lambda}) + 2\beta_{\max}\epsilon)} \quad (\text{B6})$$

(*Proof of (B5) and (B6)*): (B5) is easily derived from the large devaiiton assumptions (7) and (8). We use the fact that in a Gibbs state, the  $i$ -th smallest energy's probability  $p_i$  satisfies  $p_i = e^{-\beta(E_i - F(\beta))}$ , where  $F(\beta)$  is the Helmholtz free energy corresponding to the Gibbs state. Because of this fact, the probability  $p_{\vec{\beta}|\vec{\lambda}}^{(n)}(\vec{i})$  of  $|\vec{i}_{|\vec{\lambda}}\rangle$  in  $\rho_{\vec{\beta}|\vec{\lambda}}^{(n)}$  satisfies

$$p_{\vec{\beta}|\vec{\lambda}}^{(n)}(\vec{i}) = e^{-\sum_k \beta_k (E_{k|\lambda_k}^{(n)}(i_k) - F_{k|\lambda_k}^{(n)}(\beta_k))}. \quad (\text{B7})$$

Because of the above and (B4), the following inequalities hold:

$$e^{-\sum_k \beta_k (n\tilde{U}_k(\beta_k, \lambda_k) - F_{k|\lambda_k}^{(n)}(\beta_k)) - \beta_{\max}\epsilon} \leq p_{\vec{\beta}|\vec{\lambda}}^{(n)}(\vec{i}) \leq e^{-\sum_k \beta_k (n\tilde{U}_k(\beta_k, \lambda_k) - F_{k|\lambda_k}^{(n)}(\beta_k)) + \beta_{\max}\epsilon}. \quad (\text{B8})$$

Therefore, because of (B5) and  $\sum_k \beta_k (n\tilde{U}_k(\beta_k, \lambda_k) - F_{k|\lambda_k}^{(n)}(\beta_k)) = n\tilde{S}_S(\vec{\beta}, \vec{\lambda}) + o(n)$ , the inequalities (B6) hold for sufficiently large  $n$ .

Now, let us take two real positive numbers  $\epsilon$  and  $\epsilon'$  satisfying  $2\beta''_{\max}(\epsilon + \epsilon') < \delta'$ . Because of (B6),  $\tilde{S}_S(\vec{\beta}, \vec{\lambda}) < \tilde{S}_S(\vec{\beta}', \vec{\lambda}')$  and  $\delta' < \tilde{S}_S(\vec{\beta}', \vec{\lambda}') - \tilde{S}_S(\vec{\beta}, \vec{\lambda})$ , the following inequality holds for sufficiently large  $n$ :

$$D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} < D_{[\vec{\beta}', \vec{\lambda}', \epsilon']}^{(n)}. \quad (\text{B9})$$

Here, we number the energy pure eigenstates  $\{|\vec{i}_{|\vec{\lambda}}\rangle \otimes |\vec{\lambda}\rangle\}$  of  $H_I^{(n)}$  as  $\{|x\rangle_I\}$  satisfying the following relations:

$$|x\rangle_I \in (\text{Region} - [\vec{\beta}, \vec{\lambda}, \epsilon]) \otimes |\vec{\lambda}\rangle \Rightarrow 1 \leq x \leq D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} \quad (\text{B10})$$

$$|x\rangle_I \in (\text{Region} - [\vec{\beta}', \vec{\lambda}', \epsilon']) \otimes |\vec{\lambda}'\rangle \Rightarrow 1 + D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} \leq x \leq D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} + D_{[\vec{\beta}', \vec{\lambda}', \epsilon']}^{(n)} \quad (\text{B11})$$

Next, we give the following approximate states of  $\rho_{\vec{\beta}|\vec{\lambda}}^{(n)}$  and  $\rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)}$ :

$$\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)} := \frac{\Pi_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)} \Pi_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}}{\text{Tr}[\Pi_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} \rho_{\vec{\beta}|\vec{\lambda}}^{(n)}]} \quad (\text{B12})$$

$$\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)} := \frac{\Pi_{[\vec{\beta}', \vec{\lambda}', \epsilon']}^{(n)} \rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \Pi_{[\vec{\beta}', \vec{\lambda}', \epsilon']}^{(n)}}{\text{Tr}[\Pi_{[\vec{\beta}', \vec{\lambda}', \epsilon']}^{(n)} \rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)}]} \quad (\text{B13})$$

We also define

$$\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'| := \sum_x |f_{0,\delta'}^{(n)}(x)\rangle\langle x| (\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}\rangle\langle\vec{\lambda}|) |x\rangle\langle f_{0,\delta'}^{(n)}(x)| \quad (\text{B14})$$

where  $f_{0,\delta'}^{(n)}(x)$  is an invertible function such that

$$f_{0,\delta'}^{(n)}(x) = \begin{cases} x + D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} & (1 \leq x \leq D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}) \\ x - D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} & (1 + D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} \leq x \leq 2D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}) \\ x & (2D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)} < x) \end{cases} \quad (\text{B15})$$

Note that  $\rho_{\vec{\beta}|\vec{\lambda}}^{(n)}$  and  $\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)}$  have the same entropy spectrum  $\tilde{S}_S(\vec{\beta}, \vec{\lambda})$ , and  $\rho_{\vec{\beta}'|\vec{\lambda}'}^{(n)}$  and  $\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)}$  have the same entropy spectrum  $\tilde{S}_S(\vec{\beta}', \vec{\lambda}')$ . Also,  $\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)}$  and  $\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)}$  has the same entropy spectrum  $\tilde{S}_S(\vec{\beta}, \vec{\lambda})$ . Therefore, the entropy spectrum of  $\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)}$  is smaller than that of  $\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)}$ . Also, both of the supports of the states  $\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|$  and  $\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|$  are on the region- $[\vec{\beta}', \vec{\lambda}', \epsilon'] \otimes |\vec{\lambda}'\rangle$  and dyagonalized the same basis  $\{|x\rangle_I\}_{x=1+D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}+D_{[\vec{\beta}', \vec{\lambda}', \epsilon']}^{(n)}}^{D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}+D_{[\vec{\beta}', \vec{\lambda}', \epsilon']}^{(n)}}$  of the region- $[\vec{\beta}', \vec{\lambda}', \epsilon'] \otimes |\vec{\lambda}'\rangle$ .

Therefore, Lemma 2 guarantees that there exist the probability  $\{p_{j,\delta'}^{(n)}\}$  and the invertible functions  $\{f_{j,\delta'}^{(n)}\}$  from  $\{x\}_{x=1+D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}+D_{[\vec{\beta}', \vec{\lambda}', \epsilon']}^{(n)}}$  to  $\{x\}_{x=1+D_{[\vec{\beta}, \vec{\lambda}, \epsilon]}^{(n)}}$  such that

$$\lim_{n \rightarrow \infty} \|\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'| - \tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|\| = 0, \quad (\text{B16})$$

$$\text{supp}[\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)}] \subset \text{region} - [\vec{\beta}', \vec{\lambda}', \epsilon'], \quad (\text{B17})$$

where

$$\tilde{\rho}_{\vec{\beta}'|\vec{\lambda}'}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'| = \sum_j p_{j,\delta'}^{(n)} \sum_x |f_{j,\delta'}^{(n)}(x)\rangle\langle x| (\tilde{\rho}_{\vec{\beta}|\vec{\lambda}}^{(n)} \otimes |\vec{\lambda}'\rangle\langle\vec{\lambda}'|) |x\rangle\langle f_{j,\delta'}^{(n)}(x)|. \quad (\text{B18})$$

Now, we give  $\{q_{j,\delta'}^{(n)}, V_{j,\delta'}^{(n)}\}$  that we seek. They are defined as follows:

$$q_{j,\delta'}^{(n)} := p_{j,\delta'}^{(n)} \quad (\text{B19})$$

$$V_{j,\delta'}^{(n)} := \sum_{x,e} |g_{j,\delta'}^{(n)}(x,e)\rangle_{IEW} \langle x,e|_{IEW} \quad (\text{B20})$$

where  $|x,e\rangle_{IEW} := |x\rangle_I \otimes |e\rangle_{EW}$  and where  $g_{j,\delta'}^{(n)}(x,e)$  is an invertible function defined as

$$g_{j,\delta'}^{(n)}(x,e) := \begin{cases} \left( f_{j,\delta'}^{(n)} \circ f_{0,\delta'}^{(n)}(x), e + E_x^{(n)} - E_{f_{j,\delta'}^{(n)} \circ f_{0,\delta'}^{(n)}(x)}^{(n)} \right) & (n(\tilde{U}_W - |\Delta\tilde{U}_S| - 4h_0) \leq e \leq n(\tilde{U}_W + |\Delta\tilde{U}_S| + 4h_0)) \\ (x,e) & (\text{Otherwise}) \end{cases} \quad (\text{B21})$$

By definition, each  $V_{j,\delta'}^{(n)}$  is energy-preserving and shift-invariant in  $h_0$ -neighbourhood of  $\tilde{U}_W$ . Therefore, we only have to show that the random unitary satisfies (A3)–(A5).

We firstly show that  $\{q_{j,\delta'}^{(n)}, V_{j,\delta'}^{(n)}\}$  satisfies (A3). By definition of  $g_{j,\delta'}^{(n)}(x,e)$ , for arbitrary  $e$  satisfying  $n(\tilde{U}_W - \delta) \leq e \leq n(\tilde{U}_W + \delta)$ , the following relation holds:

$$\tilde{\rho}_{\tilde{\beta}'|\tilde{\lambda}'}^{(n)} \otimes |\tilde{\lambda}'\rangle\langle\tilde{\lambda}'| = \text{Tr}_{EW} \left[ \sum_j q_{j,\delta'}^{(n)} \sum_x V_{j,\delta'}^{(n)} (\tilde{\rho}_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}| \otimes |e\rangle\langle e|_{EW}) V_{j,\delta'}^{(n)\dagger} \right]. \quad (\text{B22})$$

Therefore, we obtain

$$\tilde{\rho}_{\tilde{\beta}'|\tilde{\lambda}'}^{(n)} \otimes |\tilde{\lambda}'\rangle\langle\tilde{\lambda}'| = \text{Tr}_{EW} \left[ \sum_j q_{j,\delta'}^{(n)} \sum_x V_{j,\delta'}^{(n)} (\tilde{\rho}_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}| \otimes \pi_{\tilde{U}_W,\delta}^{(n)}) V_{j,\delta'}^{(n)\dagger} \right]. \quad (\text{B23})$$

Because of (B5) and the gentle measurement lemma [35], the following inequalities hold for the sufficiently large  $n$ :

$$\|\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}| - \tilde{\rho}_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}|\| \leq e^{-n\alpha/2} \quad (\text{B24})$$

$$\|\rho_{\tilde{\beta}'|\tilde{\lambda}'}^{(n)} \otimes |\tilde{\lambda}'\rangle\langle\tilde{\lambda}'| - \tilde{\rho}_{\tilde{\beta}'|\tilde{\lambda}'}^{(n)} \otimes |\tilde{\lambda}'\rangle\langle\tilde{\lambda}'|\| \leq e^{-n\alpha/2} \quad (\text{B25})$$

From (B16), (B23), (B24) and (B25), we obtain

$$\lim_{n \rightarrow \infty} \left\| \text{Tr}_{EW} \left[ \sum_j q_{j,\delta'}^{(n)} \sum_x V_{j,\delta'}^{(n)} (\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}| \otimes \pi_{\tilde{U}_W,\delta}^{(n)}) V_{j,\delta'}^{(n)\dagger} \right] - \rho_{\tilde{\beta}'|\tilde{\lambda}'}^{(n)} \otimes |\tilde{\lambda}'\rangle\langle\tilde{\lambda}'| \right\| = 0. \quad (\text{B26})$$

Because of  $\sigma_{I,\delta'}^{(n)} = \text{Tr}_{EW} [\sum_j q_{j,\delta'}^{(n)} \sum_x V_{j,\delta'}^{(n)} (\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}| \otimes \pi_{\tilde{U}_W,\delta}^{(n)}) V_{j,\delta'}^{(n)\dagger}]$ , (A3) holds.

Next, we show that  $\{q_{j,\delta'}^{(n)}, V_{j,\delta'}^{(n)}\}$  satisfies (A4) and (A5). In order to show this, we firstly note that  $\sigma_{EW}^{e,(n)} := \text{Tr}_I [\sum_j q_{j,\delta'}^{(n)} \sum_x V_{j,\delta'}^{(n)} (\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)} \otimes |\tilde{\lambda}\rangle\langle\tilde{\lambda}| \otimes |e\rangle\langle e|_{EW}) V_{j,\delta'}^{(n)\dagger}]$  can be described as follows:

$$\sigma_{EW}^{e,(n)} = \sum_{e'} r_{e'} |e - n\Delta\tilde{U}_S + e'\rangle\langle e - n\Delta\tilde{U}_S + e'| \quad (\text{B27})$$

where  $r_{e'}$  is a probability distribution satisfying

$$r_{\text{main}} = \text{Tr}[\Pi_{[\tilde{\beta},\tilde{\lambda},\epsilon]}^{(n)} \rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)}] \geq 1 - e^{-n\alpha} \quad (\text{B28})$$

$$(1 - r_{\text{main}}) H\{\tilde{r}_{e'}^{(n)}\} = -(1 - r_{\text{main}}) \text{Tr}[\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)\neg} \log \rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)}] \stackrel{(a)}{=} o(n). \quad (\text{B29})$$

where

$$r_{\text{main}} := \sum_{-n(\epsilon+\epsilon') \leq e' \leq n(\epsilon+\epsilon')} r_{e'}, \quad (\text{B30})$$

$$\tilde{r}_{e'}^{(n)} := \frac{r_{e'}}{1 - r_{\text{main}}}, \quad \text{for } e' \notin (-n(\epsilon + \epsilon'), n(\epsilon + \epsilon')) \quad (\text{B31})$$

$$\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)\neg} := \frac{(1 - \Pi_{[\tilde{\beta},\tilde{\lambda},\epsilon]}^{(n)}) \rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)} (1 - \Pi_{[\tilde{\beta},\tilde{\lambda},\epsilon]}^{(n)})}{\text{Tr}[(1 - \Pi_{[\tilde{\beta},\tilde{\lambda},\epsilon]}^{(n)}) \rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)}]} \quad (\text{B32})$$

and (a) is shown by following inequality:

$$\begin{aligned} S(\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)}) &= S(r_{\text{main}}\tilde{\rho}_{\tilde{\beta}|\tilde{\lambda}}^{(n)} + (1-r_{\text{main}})\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)\neg}) \geq r_{\text{main}}S(\tilde{\rho}_{\tilde{\beta}|\tilde{\lambda}}^{(n)}) + (1-r_{\text{main}})S(\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)\neg}) \\ &\stackrel{(b)}{\geq} r_{\text{main}}(S(\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)}) - e^{-n\alpha/2} \log e^{n(\tilde{S}_S(\tilde{\beta}, \tilde{\lambda}) + 2\beta''_{\max}\epsilon)}) + (1-r_{\text{main}})S(\rho_{\tilde{\beta}|\tilde{\lambda}}^{(n)\neg}) \end{aligned} \quad (\text{B33})$$

where (b) is shown by Fannes inequality, (B5), (B6) and (B24).

Because  $\pi_{\tilde{U}_W, \delta}^{(n)}$  is described as

$$\pi_{\tilde{U}_W, \delta}^{(n)} = \frac{1}{C} \sum_{-n(\tilde{U}_W - \delta) \leq e \leq n(\tilde{U}_W + \delta)} |e\rangle\langle e|, \quad (\text{B34})$$

the state  $\sigma_{EW}^{(n)} := \text{Tr}_I[\sigma_{IEW}^{(n)}]$  is described as follows:

$$\sigma_{EW}^{(n)} = \sum_{n(\tilde{U}_W - \delta) \leq e \leq n(\tilde{U}_W + \delta)} \frac{1}{C} \sum_{e'} r_{e'} |e - n\Delta\tilde{U}_S + e'\rangle\langle e - n\Delta\tilde{U}_S + e'| \quad (\text{B35})$$

$$= \sum_{e'} r_{e'} \sum_{n(\tilde{U}_W - \delta) \leq e \leq n(\tilde{U}_W + \delta)} \frac{1}{C} |e - n\Delta\tilde{U}_S + e'\rangle\langle e - n\Delta\tilde{U}_S + e'| \quad (\text{B36})$$

$$= \sum_{e'} r_{e'} \pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)} \quad (\text{B37})$$

Therefore,

$$S(\sigma_{EW}^{(n)}) \geq \sum_{e'} r_{e'} S(\pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)}) = S(\pi_{\tilde{U}_W, \delta}^{(n)}). \quad (\text{B38})$$

Thus, in order to show (A5), we only have to show  $S(\sigma_{EW}^{(n)}) \leq S(\pi_{\tilde{U}_W, \delta}^{(n)}) + o(n)$ .

Let us show  $S(\sigma_{EW}^{(n)}) \leq S(\pi_{\tilde{U}_W, \delta}^{(n)}) + o(n)$ . With using

$$\pi_{\text{main}}^{(n)} := \frac{\sum_{-n(\epsilon + \epsilon') \leq e' \leq n(\epsilon + \epsilon')} r_{e'} \pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)}}{r_{\text{main}}}, \quad (\text{B39})$$

the state  $\sigma_{EW}^{(n)}$  is described as

$$\sigma_{EW}^{(n)} = r_{\text{main}}\pi_{\text{main}}^{(n)} + (1-r_{\text{main}}) \sum_{e' \notin (-n(\epsilon + \epsilon'), n(\epsilon + \epsilon'))} \tilde{r}_{e'}^{(n)} \pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)}. \quad (\text{B40})$$

Therefore,  $S(\sigma_{EW}^{(n)})$  satisfies

$$\begin{aligned} S(\sigma_{EW}^{(n)}) &\leq H\{r_{\text{main}}, 1-r_{\text{main}}\} + r_{\text{main}}S(\pi_{\text{main}}^{(n)}) + (1-r_{\text{main}})S\left(\sum_{e' \notin (-n(\epsilon + \epsilon'), n(\epsilon + \epsilon'))} \tilde{r}_{e'}^{(n)} \pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)}\right) \\ &\leq H\{r_{\text{main}}, 1-r_{\text{main}}\} + r_{\text{main}}S(\pi_{\text{main}}^{(n)}) + (1-r_{\text{main}})S\left(\sum_{e' \notin (-n(\epsilon + \epsilon'), n(\epsilon + \epsilon'))} \tilde{r}_{e'}^{(n)} \pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)}\right) \\ &\leq H\{r_{\text{main}}, 1-r_{\text{main}}\} + r_{\text{main}}S(\pi_{\text{main}}^{(n)}) + (1-r_{\text{main}}) \left( H\{\tilde{r}_{e'}^{(n)}\} + \sum_{e' \notin (-n(\epsilon + \epsilon'), n(\epsilon + \epsilon'))} \tilde{r}_{e'}^{(n)} S(\pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)}) \right) \\ &= H\{r_{\text{main}}, 1-r_{\text{main}}\} + r_{\text{main}}S(\pi_{\text{main}}^{(n)}) + (1-r_{\text{main}})H\{\tilde{r}_{e'}^{(n)}\} + \sum_{e' \notin (-n(\epsilon + \epsilon'), n(\epsilon + \epsilon'))} \tilde{r}_{e'}^{(n)} S(\pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)}) \\ &\leq \log 2 + r_{\text{main}}S(\pi_{\text{main}}^{(n)}) + o(n) + \sum_{e' \notin (-n(\epsilon + \epsilon'), n(\epsilon + \epsilon'))} \tilde{r}_{e'}^{(n)} S(\pi_{\tilde{U}_W - \Delta\tilde{U}_S + \frac{e'}{n}, \delta}^{(n)}) \\ &= \log 2 + r_{\text{main}}S(\pi_{\text{main}}^{(n)}) + (1-r_{\text{main}})S(\pi_{\tilde{U}_W, \delta}^{(n)}) + o(n). \end{aligned} \quad (\text{B41})$$

Let us decompose  $\pi_{\text{main}}$  as follows:

$$\pi_{\text{main}} = r_A \pi_A + r_B \pi_B \quad (\text{B42})$$

where

$$r_A := \text{Tr}[\Pi_A \pi_{\text{main}}] \quad (\text{B43})$$

$$\pi_A := \frac{\Pi_A \pi_{\text{main}} \Pi_A}{r_A} \quad (\text{B44})$$

$$\Pi_A := \sum_{n(\tilde{U}_W - \Delta \tilde{U}_S - \delta + \epsilon + \epsilon') \leq e \leq n(\tilde{U}_W - \Delta \tilde{U}_S + \delta - \epsilon - \epsilon')} |e\rangle\langle e| \quad (\text{B45})$$

$$r_B := \text{Tr}[\Pi_B \pi_{\text{main}}] \quad (\text{B46})$$

$$\pi_B := \frac{\Pi_B \pi_{\text{main}} \Pi_B}{r_B} \quad (\text{B47})$$

$$\Pi_B := \sum_{n(\tilde{U}_W - \Delta \tilde{U}_S - \delta - \epsilon - \epsilon') \leq e \leq n(\tilde{U}_W - \Delta \tilde{U}_S + \delta + \epsilon + \epsilon')} |e\rangle\langle e| - \Pi_A \quad (\text{B48})$$

Because the support of  $\pi_{\text{main}}$  is on the subspace projected by the projection  $\Pi_A + \Pi_B$ , the equality  $r_A + r_B = 1$  holds. Because an arbitrary quantum state  $\rho$  has entropy lower than the maximally mixed state on the support of  $\rho$ , we obtain

$$S(\pi_A) \leq S(\pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta - \epsilon - \epsilon'}^{(n)}) \quad (\text{B49})$$

$$S(\pi_B) \leq S\left(\frac{\pi_{\tilde{U}_W - \Delta \tilde{U}_S - \delta, \epsilon + \epsilon'}^{(n)} + \pi_{\tilde{U}_W - \Delta \tilde{U}_S + \delta, \epsilon + \epsilon'}^{(n)}}{2}\right). \quad (\text{B50})$$

Similarly, because the support of  $\pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta - \epsilon - \epsilon'}^{(n)}$  is included in the support of  $\pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)}$ , we obtain

$$S(\pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta - \epsilon - \epsilon'}^{(n)}) \leq S(\pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)}) = S(\pi_{\tilde{U}_W, \delta}^{(n)}). \quad (\text{B51})$$

Because of  $2\beta''_{\text{max}}(\epsilon + \epsilon') < \delta' < 2\beta''_{\text{max}}\delta$ , we obtain

$$S\left(\frac{\pi_{\tilde{U}_W - \Delta \tilde{U}_S - \delta, \epsilon + \epsilon'}^{(n)} + \pi_{\tilde{U}_W - \Delta \tilde{U}_S + \delta, \epsilon + \epsilon'}^{(n)}}{2}\right) \leq H\left\{\frac{1}{2}, \frac{1}{2}\right\} + \frac{1}{2}S(\pi_{\tilde{U}_W - \Delta \tilde{U}_S - \delta, \epsilon + \epsilon'}^{(n)}) + \frac{1}{2}S(\pi_{\tilde{U}_W - \Delta \tilde{U}_S + \delta, \epsilon + \epsilon'}^{(n)}) \quad (\text{B52})$$

$$\begin{aligned} &\leq H\left\{\frac{1}{2}, \frac{1}{2}\right\} + \frac{1}{2}S(\pi_{\tilde{U}_W - \Delta \tilde{U}_S - \delta, \delta}^{(n)}) + \frac{1}{2}S(\pi_{\tilde{U}_W - \Delta \tilde{U}_S + \delta, \delta}^{(n)}) \\ &= \log 2 + S(\pi_{\tilde{U}_W, \delta}^{(n)}). \end{aligned} \quad (\text{B53})$$

Therefore, we obtain

$$S(\pi_{\text{main}}) = H\{r_A, r_B\} + r_A S(\pi_A) + r_B S(\pi_B) \leq 2 \log 2 + S(\pi_{\tilde{U}_W, \delta}^{(n)}) \quad (\text{B54})$$

Therefore,

$$S(\sigma_{E_W}^{(n)}) \leq 3 \log 2 + S(\pi_{\tilde{U}_W, \delta}^{(n)}) + o(n) \quad (\text{B55})$$

holds. Thus,  $\{q_{j, \delta'}^{(n)}, V_{j, \delta'}^{(n)}\}$  satisfies (A5).

Finally, we show that  $\{q_{j, \delta'}^{(n)}, V_{j, \delta'}^{(n)}\}$  satisfies (A4). Note that  $\pi_A$  is the maximally mixed state on the subspace projected by  $\Pi_A$ , and that  $\pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)}$  is described as

$$\pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)} = \frac{\delta - (\epsilon + \epsilon')}{\delta} \pi_A + \frac{\epsilon + \epsilon'}{\delta} \pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta, B}^{(n)}, \quad (\text{B56})$$

where

$$\pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta, B}^{(n)} := \frac{\Pi_B \pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)} \Pi_B}{\text{Tr}[\Pi_B \pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)}]}. \quad (\text{B57})$$

Therefore, we obtain

$$\|\pi_{\text{main}} - \pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)}\| \leq 2 \left| r_A - \frac{\delta - (\epsilon + \epsilon')}{\delta} \right| + \left\| r_B \pi_B - \frac{\epsilon + \epsilon'}{\delta} \pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)} \right\| \leq 2 \left| r_A - \frac{\delta - (\epsilon + \epsilon')}{\delta} \right| + 2(r_B + \frac{\epsilon + \epsilon'}{\delta}). \quad (\text{B58})$$

Because of  $r_A \geq \frac{\delta - \epsilon - \epsilon'}{\delta}$ , the inequality  $r_B \leq \frac{\epsilon + \epsilon'}{\delta} \leq \frac{\delta'}{2\beta''_{\max}\delta}$  holds. Therefore,

$$\|\pi_{\text{main}} - \pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)}\| \leq \frac{4\delta'}{\beta''_{\max}\delta} \quad (\text{B59})$$

Because of (B28) and (B40), we obtain

$$\|\sigma_{E_W}^{(n)} - \pi_{\text{main}}\| \leq O(e^{-n\alpha}). \quad (\text{B60})$$

Hence,

$$\lim_{n \rightarrow \infty} \|\sigma_{E_W}^{(n)} - \pi_{\tilde{U}_W - \Delta \tilde{U}_S, \delta}^{(n)}\| \leq \frac{4\delta'}{\beta''_{\max}\delta} \quad (\text{B61})$$

Therefore,  $\{q_{j, \delta'}^{(n)}, V_{j, \delta'}^{(n)}\}$  satisfies (A4). ■

### Appendix C: Proof of Corollaries 3 and 4

**Proof:** Let us take  $\beta''_M$  and  $\tilde{W}_M''$  arbitrarily. Then, the adiabatic work extraction  $(\beta, \lambda_S) \times (\beta, \lambda_B) \rightarrow_{\text{ad:}\tilde{W}_M''} (\beta, \lambda'_S) \times (\beta''_M, \lambda_B)$  is possible if and only if the following two expressions hold:

$$\tilde{S}_S(\beta, \lambda_S) + \tilde{S}_{B_M}(\beta, \lambda_B) \leq \tilde{S}_S(\beta, \lambda'_S) + \tilde{S}_{B_M}(\beta''_M, \lambda_B), \quad (\text{C1})$$

$$\tilde{W}_M'' = -\Delta \tilde{U}_S + \tilde{Q}_M'', \quad (\text{C2})$$

where  $\tilde{Q}_M'' := \tilde{U}_{B_M}(\beta, \lambda_B) - \tilde{U}_{B_M}(\beta''_M, \lambda_B)$ . Substituting (12) into (C1) and (C2), we obtain the following necessary and sufficient condition of the possibility of the adiabatic work extraction  $(\beta, \lambda_S) \times (\beta, \lambda_B) \rightarrow_{\text{ad:}\tilde{W}_M''} (\beta, \lambda'_S) \times (\beta''_M, \lambda_B)$ :

$$\tilde{W}_M'' \leq -\Delta \tilde{F}_S - \frac{\int_{\tilde{U}_{B_M}(\beta, \lambda_B)}^{\tilde{U}_{B_M}(\beta, \lambda_B) - X_M} \beta_M(u) du - \beta \int_{\tilde{U}_{B_M}(\beta, \lambda_B)}^{\tilde{U}_{B_M}(\beta, \lambda_B) - X_M} du}{\beta} \quad (\text{C3})$$

$$\tilde{W}_M'' = -\Delta \tilde{U}_S + \tilde{Q}_M'', \quad (\text{C4})$$

where  $\beta_M(u)$  is the reverse function of  $\tilde{U}_{B_M}(\beta, \lambda_B)$  (Note that  $\lambda_B$  is fixed.), and where

$$X_M := \tilde{W}_M'' + \Delta \tilde{U}_S. \quad (\text{C5})$$

Note that  $\tilde{X}$  is not equal to  $\tilde{Q}_M''$  in general, although they are equal to each others when (C4) holds. Moreover, the second term of the righthand side of (C3) satisfies

$$\left| \frac{\int_{\tilde{U}_{B_M}(\beta, \lambda_B)}^{\tilde{U}_{B_M}(\beta, \lambda_B) - X_M} \beta_M(u) du - \beta \int_{\tilde{U}_{B_M}(\beta, \lambda_B)}^{\tilde{U}_{B_M}(\beta, \lambda_B) - X_M} du}{\beta} \right| \leq \frac{|\beta - \beta''_M|}{\beta} |X_M|. \quad (\text{C6})$$

Now we have completed the preparation. We firstly prove Corollary 3. When  $\tilde{W} \leq -\Delta \tilde{F}_S$  holds, we define  $\tilde{W}_M$  and  $\tilde{\beta}'_M$  as follows:

$$\tilde{W}_M := \tilde{W} - \frac{1}{\sqrt{M}}, \quad (\text{C7})$$

$$\tilde{\beta}'_M := \beta_M(\tilde{U}_{B_M}(\beta, \lambda_M) - \tilde{W}_M - \Delta \tilde{U}_S) \quad (\text{C8})$$



Clearly,  $\tilde{W}_M$  satisfies  $\tilde{W}_M = \tilde{W} + O(1/\sqrt{M})$ . Because the specific heat  $B_M$  is proportional to  $M$ ,  $\tilde{\beta}'_M = \beta + O(1/M)$  holds. Therefore, because of (C6), the order of the second term of the righthand side of (C4) is  $O(1/M)$ . Hence, when we define the functions  $\tilde{W}_M$  and  $\beta'_M$  as  $\tilde{W}_M''$  and  $\beta'_M''$ ,  $\tilde{W}_M$  and  $\beta'_M$  satisfy (C3) and (C4) for sufficiently large  $M$ . Therefore, there exist  $\tilde{W}_M$  and  $\beta'_M$  satisfying  $\lim_{M \rightarrow \infty} \tilde{W}_M = \tilde{W}$  and  $\lim_{M \rightarrow \infty} \beta'_M = \beta$ , and the adiabatic work extraction  $(\beta, \lambda_S) \times (\beta, \lambda_B) \rightarrow_{\text{ad:}\tilde{W}_M} (\beta, \lambda'_S) \times (\beta'_M, \lambda_B)$  is possible for sufficiently large  $M$ . Therefore, when  $\tilde{W} \leq -\Delta\tilde{F}_S$  holds, the isothermal work extraction  $(\beta, \lambda_S) \rightarrow_{\text{is:}\tilde{W}} (\beta, \lambda'_S)$  is possible.

Next, let us show that when  $\tilde{W} > -\Delta\tilde{F}_S$  holds, the isothermal work extraction  $(\beta, \lambda_S) \rightarrow_{\text{is:}\tilde{W}} (\beta, \lambda'_S)$  is impossible. We define a number  $\delta > 0$  as  $\delta := \tilde{W} + \Delta\tilde{F}_S$ . We also take arbitrary functions  $\tilde{W}_M$  and  $\beta'_M$  satisfying  $\tilde{W}_M = \tilde{W} + o(1)$  and  $\beta'_M = \beta + o(1)$ . Substituting  $\tilde{W}_M''$  and  $\beta'_M''$  for (C3) and (C4), we obtain that the order of the second term of the righthand side of (C4) is  $o(1)$ . On the other hand, the lefthand side of (C4) is equal to  $-\Delta F_S + \delta + o(1)$ . Therefore, (C4) is invalid for sufficiently large  $M$ . Therefore, no matter what  $\tilde{W}_M$  and  $\beta'_M$  satisfying  $\tilde{W}_M = \tilde{W} + o(1)$  and  $\beta'_M = \beta + o(1)$  are used, the adiabatic work extraction  $(\beta, \lambda_S) \times (\beta, \lambda_B) \rightarrow_{\text{ad:}\tilde{W}_M} (\beta, \lambda'_S) \times (\beta'_M, \lambda_B)$  is impossible for sufficiently large  $M$ . Therefore, when  $\tilde{W} > -\Delta\tilde{F}_S$  holds, the isothermal work extraction  $(\beta, \lambda_S) \rightarrow_{\text{is:}\tilde{W}} (\beta, \lambda'_S)$  is impossible.

Finally, let us show Corollary 4. When  $(\beta, \lambda_S) \rightarrow_{\text{is:}\tilde{W}} (\beta, \lambda'_S)$  is possible, there exist two functions  $\tilde{W}_M$  and  $\beta'_M$  satisfying  $\tilde{W}_M = \tilde{W} + o(1)$  and  $\beta'_M = \beta + o(1)$ , and the adiabatic work extraction  $(\beta, \lambda_S) \times (\beta, \lambda_B) \rightarrow_{\text{ad:}\tilde{W}_M} (\beta, \lambda'_S) \times (\beta'_M, \lambda_B)$  is possible. Let us take arbitrary set of  $\tilde{W}_M$  and  $\beta'_M$  satisfying  $\tilde{W}_M = \tilde{W} + o(1)$  and  $\beta'_M = \beta + o(1)$  and satisfying that the adiabatic work extraction  $(\beta, \lambda_S) \times (\beta, \lambda_B) \rightarrow_{\text{ad:}\tilde{W}_M} (\beta, \lambda'_S) \times (\beta'_M, \lambda_B)$  is possible. Then, when we substitute  $\tilde{W}_M$  and  $\beta'_M$  for  $\tilde{W}_M''$  and  $\beta'_M''$  in (C4), the relation (C4) holds. Therefore, Corollary 4 holds. ■